DOCUMENT RESUME

ED 171 743 TM 008 772

AUTHOR	Olson, Paul
TITLE	A View of Power: Four Essays on the National Assessment of Educational Progress.
INSTITUTION	North Dakota Study Group on Evaluation, Grand Forks.
SPONS AGENCY	Rockefeller Bros. Fund, New York, N.Y.
PUB DATE	Dec 76
NOTE	44p.
AVAILABLE FROM	Center for Teaching and Learning, University of North Dakota, Grand Forks, North Dakota 58202 ($2.00)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Community Control; Culture Free Tests; *Educational Assessment; Educational Objectives; Elementary Secondary Education; Group Norms; Literature Appreciation; *Middle Class Values; *National Competency Tests; *National Norms; Reading Tests; School District Autonomy; *Test Bias; *Testing Problems; Testing Programs; Writing Skills
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

In a series of essays, the author analyzes and comments on the perceived weaknesses of the National Assessment of Educational Progress (NAEP) testing program. The basis of his criticism is that the NAEP has attempted to establish national norms, thereby disregarding the pluralistic nature of American society and the historical basis for community control of education; that NAEP has offered a facade of national consensus by ignoring input from many minority groups; and that local and less powerful groups have been asked to assess themselves in a context established by more powerful establishment groups. The author reviews the NAEP's literature, writing and reading components, and finds them oriented toward Western culture, unrelated to a meaningful context for ethnic groups and geared to skills which are not environment-specific and are dictated by the need to satisfy a national norm. The author recommends local context-specific assessment programs by which the community can determine whether it is developing a solid educational program and preparing children to function in the national, as well as local, society. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

North Dakota Study Group on Evaluation

ED171743

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paul Olson

A VIEW OF POWER: FOUR ESSAYS ON THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

TM008 772

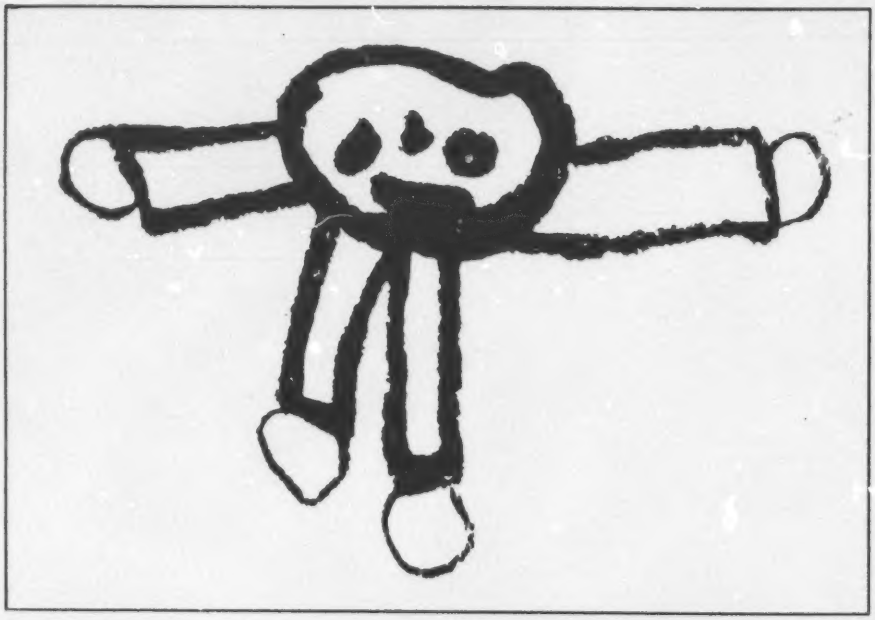
"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

V. Perrone

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

2

North Dakota Study Group on Evaluation



Paul Olson

A VIEW OF POWER: FOUR ESSAYS ON THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

University of North Dakota
Grand Forks, ND 58202
December 1976

Copyright © 1976 by Paul Olson

First published in 1976

North Dakota Study Group
on Evaluation, c/o Vito Perrone,
Center for Teaching & Learning
University of North Dakota
Grand Forks, N.D. 58201

Library of Congress Catalogue
Card Number: 76-57117

Printed by University of
North Dakota Press

A grant from the Rockefeller Brothers Fund
makes possible publication of this series

Editor: Arthur Tobier

In November 1972, educators from several parts of the United States met at the University of North Dakota to discuss some common concerns about the narrow accountability ethos that had begun to dominate schools and to share what many believed to be more sensible means of both documenting and assessing children's learning. Subsequent meetings, much sharing of evaluation information, and financial and moral support from the Rockefeller Brothers Fund have all contributed to keeping together what is now called the North Dakota Study Group on Evaluation. A major goal of the Study Group, beyond support for individual participants and programs, is to provide materials for teachers, parents, school administrators and governmental decision-makers (within State Education Agencies and the U.S. Office of Education) that might encourage re-examination of a range of evaluation issues and perspectives about schools and schooling.

Towards this end, the Study Group has initiated a continuing series of monographs, of which this paper is one. Over time, the series will include material on, among other things, children's thinking, children's language, teacher support systems, inservice training, the school's relationship to the larger community. The intent is that these papers be taken not as final statements--a new ideology, but as working papers, written by people who are acting on, not just thinking about, these problems, whose implications need an active and considered response.

Vito Perrone, Dean
Center for Teaching & Learning,
University of North Dakota

Contents

1 The Courts vs. the Professionals: A View of Power 1

2 The Night Chant vs. The Village Blacksmith: A View of Literature 13

3 Rap Brown vs. Betsy Ross: A View of Writing 16

4 The National Assessment of Reading: Horsepower Without Horse Sense Is Fatal 27

5 Toward an Alternative to National Assessment 34

1

The Courts vs. the Professionals: A View of Power

I have been asked to write about the National Assessment of Educational Progress. Though I am not a specialist in assessing educational progress, I am a citizen. I have children. I have taught in schools and colleges in all kinds of cultural contexts, and I have an interest in how people, communities, and school systems are valued by other people or institutions. In particular, I am interested in how the powerless or different in America are valued, and encouraged to value themselves, by the powerful and *well-respected* people of our world. By the powerless or different, I mean those persons who would, under law, be clients of the ESEA Title I program, the Bilingual Education Act, the Indian Education Act, and other similar programs. I also mean those persons who would not be comfortable at T.S. Eliot's *Cocktail Party*. My thesis is that the National Assessment represents:

- an assessment based on putative national norms in a period when the courts are saying that local norms and parent or community concerns are crucial;
- an assessment that offers the appearance of a national consensus with respect to what education should do only by virtue of leaving out many of the concerned parties;
- an assessment that, at best, asks powerless communities to assess themselves in terms provided by the powerful.

A CONTEXT FOR UNDERSTANDING THE ASSESSMENT

The *National Norms* on which the Assessment is based are reflections of national textbook culture, at least as regards the social and political sciences; they *do not* arise from any discovery that there are universally accepted intellectual perspectives in these areas or universally accepted patterns of social or artistic behavior or appreciation.

Nor is the notion of a National Assessment of Educational Progress a new one. It is at least a century old, and should perhaps be celebrated in the Bicentennial. The original Department of Education bill, prepared as part of

Paul Olson is professor of English at the University of Nebraska in Lincoln.

the Reconstruction legislation in 1867, included a plan "to enforce education, without regard to race or color, upon the population of all such states as shall fall below a standard to be established by Congress."[1] This language was rejected; had it been passed at the time of the creation of the Department of Education, the National Assessment would now be more than 100 years old, and we would have a clearer understanding of its effects. Instead, however, Congress chose much milder language calling for a Department of Education that would be a service organization to the people of the United States:

> Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, that there shall be established, at the city of Washington, a Department of Education, for the purpose of collecting such statistics and facts as shall show the condition and progress of education in the several states and Territories, and of diffusing such information respecting the organization and management of schools and school systems, and methods of teaching, as shall aid the people of the United States in the establishment and maintenance of efficient school systems, and otherwise promote the cause of education throughout the country.[2]

1. Compare Donald R. Warren, *To Enforce Education: A History of the Founding Years of the United States Office of Education* (Detroit: Wayne State University Press, 1974), pp. 78, 204-05, and 58-97 passim.

2. Warren, *To Enforce Education*, pp. 204-05.

It is clear from the debate record that *the people of the United States* means *the people in their natural local communities* and not the nation-state.

The thinking that led the Congress to reject a national standard and its enforcement is still vital. It is reflected in the Supreme Court's statement in the recent Detroit desegregation decision, which, for better or for worse, affirms that community control of education is not only an old tradition in this country but confers significant educational benefits in that it permits the development of an organic relation between school and home and the creation of experiment, competition, and new directions that improve education:

> No single tradition in public education is more deeply rooted than local control over the operation of schools; local autonomy has long been thought essential both to the maintenance of community concern and support for public schools and to quality of the educational process. ... Thus, in *San Antonio School District vs. Rodriguez*, 411 U.S. 1, 50, we observed that local control over the educational process affords citizens an opportunity to participate in decision-making, permits the structuring of school programs to fit local needs, and encourages "experimentation, innovation and a healthy competition for educational excellence."[3]

3. *Milliken v. Bradley*, July 1974

This thinking figures prominently, too, in the various federal and circuit cases that interpret education as an extension of child rearing, such as *Pierce v. Society of Sisters, Meyer v. Nebraska, Farrington v. Tokushige,* and *Wisconsin v. Yoder.*[4]

Respect for the local, unique culture as the prime determinant in the construction of educational processes is also reflected in the several recent court decisions giving people the right to education in their own language (*Aspira, Lau,* and *Portales*). It speaks through the various decisions which require that teachers be educated and licensed to fit local job conditions and job descriptions (*Mercado and Chance v. New York City Board of Examiners; Walston v. Nansemond County*). I feel that something of the spirit of the original effort to avoid a national standard speaks in Justice William O. Douglas's opinion in *Defunis v. Odegaard,* a case that the rest of the Court regarded as moot. Douglas wrote of the Law School Admissions Tests:

> Insofar as the LSAT tests reflect the dimensions and orientation of the Organization Man, they do a disservice to minorities. I personally know that admissions tests were once used to eliminate Jews. How many other minorities they aim at I do not know. ... The invention of substitute tests might be made to get a measure of an applicant's cultural background, perception, ability to analyze, *and his or her relation to groups.*[5] (Emphasis added)

In looking at the LSATs, Douglas affirms that the candidate's relation in his local group is important in norm construction and that different groups may look to different thinkers: The Native American of Washington state to Chief Seattle or Chief Joseph rather than Adam Smith or Marx, black groups to Sekou Toure or Senghor rather than to Smith or Marx, and so on.

In short, in a variety of cases, and with some consistency, the courts have seen education as a local matter--an extension of child rearing--requiring different approaches, norms, and curricula for different people.

There is, I think, an alternative to the Court view, which, for the want of a better word, I shall call the *professional* view. By *professional* in this context, I do not mean the competent view, or the view that helps children learn most and best. Rather, I mean the view that has come into fashion as defining what spokesmen for professional education agencies are expected to defend when they address the public--the view to which the person entering the profession tends to be tribalized. This view is formulated well by Stephen Bailey of the American Council on Education:

> The language (of local control) falls pleasingly on the ears of local school board members, superinten-

4. *Pierce vs. Society of Sisters,* 268 U.S. 510, 534-35; *Meyer v. Nebraska,* 262 U.S. 390, 401-02; *Farrington v. Tokushige,* 273 U.S. 284, 293-94; and *Wisconsin v. Yoder,* 406 U.S. 205.

5. William O. Douglas, opinion in *DeFunis v. Odegaard,* reproduced in *Study Commission* (May 1974), supplement.

dents, teachers and parents; and it may well be a barrier to arbitrariness at higher levels. But the term "local-control"--powerful as it is as a political shibboleth--flies in the face of the fiscal and administrative realities of state and federal grants in aid and the standardizing effect of professionalism upon public education across the land. This struggle between shibboleth and reality is one of the political anomalies of educational finance. Part of the political tension which surrounds and infuses contemporary educational finance controversies derives from the fact that in a highly interdependent, technological world, the myth of local control of educational polity is increasingly unrealistic.[6]

6. Joel S. Berke and Michael W. Kirst, *Federal Aid to Education: Who Benefits? Who Governs?* (Lexington: Lexington Books, 1972), pp. 22-23.

I wish to argue that the National Assessment of Educational Progress, in its power base, conception, creation, and execution, runs against what I have described as the Court view and supports what I have called the professional view. Fundamental to the professional view, I think, is that education is an entity and that it is the function of national programs to clarify what that entity is, to develop a *delivery system* to deliver that entity to children, and to certify that the delivery has, in fact, been made.

A BRIEF HISTORY

One history of the origin of the National Assessment of Educational Progress appears in a study of testing by the Youth Project, a group related to Ralph Nader's organization.[7] According to that study, the NAEP program was itself the product of no local educational impulse or public debate. It had its source in a committee established by the Carnegie Corporation in 1964 to examine the feasibility of developing some mechanism for assessing the educational programs and learning levels in the United States.

7. Youth Project, *The Project on Educational Testing: Interim Report*, pp. 123-29.

By 1965, when it was provided a grant of $260,000, the committee had already developed descriptions of necessary and allegedly unique testing instruments and had got the Psychological Corporation, Science Research Associates, the American Institute for Research, and Educational Testing Service to help develop the instruments. The Ford Foundation put $496,000 into the effort in 1966, and Carnegie put in $640,000 more the following year. The then-Commissioner Harold Howe, who is now with the Ford Foundation, first involved the Office of Education directly in 1968--with a grant of $1 million. When opposition developed from the American Association of School Administrators, a former president of AASA was made chairman of the National Assessment Committee. An *independent* supervisor of the committee--the Education Commission of the States, a Carnegie-supported coalition--took over

supervision of the National Assessment for the Carnegie Corporation. The executive director of ECS, Wendell Pierce, had, in turn, served on the ETS Board of Directors from 1965 to 1969.* In testimony before the Mondale Committee in 1972, Francis Keppel, after presenting a very attractive account of the disappearance of opposition to the NAEP among school administrators and state people, remarked as follows:

> There is an extraordinary hopeful possibility that out of this movement we can develop measures by the school--the program within the school building--which will make it possible...to rifle-shoot direct funds to improve the performance within a school building.
>
> I am making a contrast here between the school system as a whole--all the primary, junior high, and high schools, treated as a unit--because the important data on equal educational opportunity gets lost in that aggregate. It would seem to me essential that we disaggregate it; get the unit of measure down to the school itself, the place where the individual in charge can be held more responsible, in my judgment, than the superintendent.
>
> On the other hand, we would not, by these techniques, be overburdening the individual children with a whale of a lot of measures. A pretty good case can be made that too much testing is not good for children. By sampling technique, it is, as far as I can see, possible to reduce the amount of testing and still give the responsible authorities like yourself the data that you need for public policy....
>
> ....When the Congress and the executive branch are satisfied that the technical problems are under control, the Office of Education could then both collect data on such a basis and make public reports. Neither is possible today.
>
> You will have noted that this testimony puts special emphasis on the need of assessing what the institution, the particular school, is accomplishing, rather than on a school system as a whole or on how the individual learner is performing. Presumably, in years to come, this committee and the Congress will want to have aggregate data in order to understand the extent of variation from the goal of equal educational opportunity as expressed in measures of what pupils have learned. A child who cannot read or cipher up to a minimum standard cannot take advantage of educational opportunities. But for the national purpose, if remedial action is to be effective, this information has to be applied to the basic management unit--the individual school--where something can be done and where responsibility can be lodged.[8]

*This information all comes from the Youth Project study (pp. 123-129); it also shows that eight of eleven members of the original exploratory committee (Ralph Tyler, John J. Corson, Devereux C. Josephs, Roy E. Larsen, Lloyd W. Morrisett, Katherine McBride, Melvin Barnes, and Paul Reinert) had had close relationships--many of them long standing--with both the Carnegie Corporation and ETS. In 1967, the AASA recommended that its members refuse to cooperate. The Carnegie Corporation expanded the committee to include several new organizations; named George Brain, former head of AASA, chairman of the new Committee on Assessing the Progress of Education, to succeed Ralph Tyler; and placed the committee with the Education Commission of the States. The Youth Project argues that "the organization which had criticized the program in 1966 had been coopted" by the end of 1968.

8. Francis Keppel, "Statement," *Hearings Before the Select Committee on Equal Opportunity, Part 22: Education Information*, pp. 10950-52.

Thus Keppel argued for the usefulness of the National Assessment not only as a kind of graph of where we are educationally in the nation, but as a different kind of local control and local management. Keppel also argued for the Office of Education using NAEP nationally to "both collect data" and "make public reports"--that is, in national management. (Keppel's father, by the way, was president of Carnegie from 1923 to 1942, and Keppel himself became a trustee of the corporation in 1970. Carnegie has, furthermore, pushed fairly consistently for a national and a standardized management of education.) If I understand Keppel's testimony correctly, local judgments would be made using national tools. Once an outside group has established what the *accounting and describing* procedures are to be, the social order has been changed, no matter what the judgments are that emerge from examining the accounts.[9]

9. Warren, *To Enforce Education.*

DEVELOPING A WIDER, MORE PERSUASIVE CLIENTELE

If the National Assessment originally came out of a fairly limited circle of consultation, its development was designed to make it have a wider, more persuasive clientele.

After the *exercises* (the National Assessment's substitute for *test items*) were developed, they were reviewed by the following *lay* groups: AAUW, AFL-CIO, students in colleges of education, NAACP, National Citizens Committee of Christians and Jews, National Congress of Parents and Teachers, state PTAs, a range of school board associates, and the U.S. Chamber of Commerce Education Committee.[10]

These lay groups were asked three questions: (1) would you object to having the exercise used with your child? (2) would any important group object to the use of the exercise? and (3) is the exercise "sufficiently important to risk offending certain segments of the population?"[11] Lay people were *not* asked whether the items measured the goals of education as conceptualized in their local community or whether they measured, in any sense, the capacity to function in an expressive or vocational or any other competent, adult role in their community. Lay participants were apparently asked only if they found something offensive in the exercises, or if they thought that some other competent group in their community would. (How the lay participants were to find out if other people would find an item offensive is difficult to know.)

10. Carmen J. Finley, *The National Assessment Approach to Exercise Development* (Denver: National Assessment of Educational Progress, 1970), p. 38. This is an official NAEP publication.

11. Ibid., p. 36.

The lay participation group appears to have been predominantly middle class and predominantly white. It seems not to have included those stubborn, brilliant, determined cultural groups that have, year after year in the last decade, raised the right questions in the courts about their right to have education conducted according to the norms of their own community and assessed in terms of them: the Mennonites in Wisconsin; the Nation of Islam; the Puerto Ricans in the Aspira organization; the sorts of organiza-

tions that were the plaintiffs in the *Lau, Portales,* and *San Felipe del Rio* cases; and the Appalachian groups that have tried to achieve culture-based education in Appalachia.

National Assessment items were later assessed for *subject matter* validity by people nominated by the International Reading Association, the National Council of Teachers of English, the National Council of Teachers of Mathematics, the National Association of Industrial Teacher Educators, the National Science Teachers Association, the American Industrial Arts Association, the American Vocational Association, the National Art Education Association, the National Council for the Social Sciences, the American Historical Association, the Music Educators National Conference, and the National Association of Schools of Music.[12] The professional societies' nominated reviewers were asked to find out if the item: (1) sampled the objective indicated; (2) indicated the correct answer to an exercise; and (3) had any ambiguities or other flaws. Reviewers were also asked to estimate how many people of various ages would answer the items correctly.

12. Ibid., pp. 47-48.

Again, it is clear that the picture of *professionalism* represented by the professional societies asked to nominate reviewers is not catholic. Why only the predominantly white educational associations and not other groups, such as the College English Association, the Indian Historical Society, or the educators who publish in *Quinto Sol* publications or in the *Black Scholar?* Why the *applied* professional societies and not the theory-oriented ones, such as the American Institute for the Biological Sciences, the Linguistic Society of America, the American Physical Association, or the Mathematical Association of America? It is, I think, because a *national assessment,* almost by definition, seeks to homogenize things. Unique culture-bearers do not get in, nor do the uniquely gifted. One doubts that the Noam Chomskys of the world were asked to review items having a linguistic or psycholinguistic content (writing, reading, literature).

The court decisions in *Lau, Portales,* and *Aspira,* which assert the right to be educated in one's own language, are particularly significant to the National Assessment efforts. All the NAEP schedules that I have reviewed use English as their only idiom. Perhaps significant to the reading examination is recent research showing the English-speaking children read best that material which uses the idiom, landscape, iconography, and stories that are used in their own area and culture.*

*This was first suggested by Paulo Freire and has been confirmed by Yetta and Ken Goodman in their researches on Appalachia. I am at a loss to know what earthly use is served by reporting reading scores or any other Assessment [illegible]es by race, unless genetic assumptions are involved. Present linguistic research might suggest classification of reading scores in terms of first language of child or school; regional or socioeconomic dialect (e.g., children who speak Gullah dialect); literacy and primary vocabulary of parents; and availability of books in the home.

The *exercise* construction process that this sort of administrative arrangement leads to may be characterized by Lawrence Freeman's description of the National Assessment of Musical Performance,[13] which was written for a preliminary version of the final report of the U.S.O.E. Study Commission on Undergraduate Education and the Education of Teachers:

13. Frank Rivas, *The First National Assessment of Musical Performance* (Denver: National Assessment of Educational Progress, 1974).

This assessment--supported by a USOE grant--was constructed in the following way. Exercises "were developed to measure the extent to which young people are achieving the objectives compiled by music professionals in 1965." The resultant test exercises include singing familiar songs, e.g., "America" and "Are You Sleeping"; repeating unfamiliar musical material; improvising; performing from notation; and performing a prepared piece. Precise criteria for judging performance were then developed. For example, in singing "America," "pitch" was judged unacceptable if a pitch was closer to the next half step than to the right original pitch; four or more pitch errors led to classifying a response as unacceptable. In improvising melody, an acceptable response had to begin "within two measures of the end of the stimulus, must not have deviated in tempo by more than 10 per cent and must have not contained more than two unidentifiable pitches." And in sight-singing "a pitch was considered to be incorrect if it was closer to the next half step than to the right pitch. Three pitch errors and one change of key were allowed in an acceptable performance. However, if one of the major second intervals were maintained consistently, the other interval may have been sung at a minor second interval without causing the responses to be scored unacceptable."

...Nowhere in the materials that have been examined is there an explicit statement about the social and cultural function of music. In the performance test referred to, performance of music is apparently viewed as an individual, not a group act; performance was required apparently in a situation in which the individual was isolated from situations which might inspire enjoyable or effective exercises in music. Music performance is apparently thought of as a solo performance in Carnegie Hall before silent, critical strangers. The instances of familiar music, "America" and "Are You Sleeping," are familiar "school" pieces, not pieces that students would enjoy spontaneously rendering. And finally the performance of music is seen as an extremely technical process with the standards derived almost exclusively from Western European "high culture" music and Western European conceptions of pitch, tonality, harmony, and performance-timing.....

...And the notion of "music" implicit in the assessment is apparently non-American, with little recognition that the local sources of inspiration for American "classical" music--extensions of Western European music--have been primarily indigenous

> black and regional cultural music. If the notions of music implicit in the assessment are those out of which teaching is supposed to arise, one can hypothesize the decimation of those indigenous sources of inspiration.[14]

Freeman's rhetoric may be a little flamboyant, but his perception is, I think, essentially accurate. Somewhere in *Notes Toward a Definition of Culture*, T.S. Eliot remarks on the importance to English life of the survival of Welsh language and culture:

> For the transmission of a culture--a peculiar way of thinking, feeling, and behaving--and for its maintenance, there is no significant safeguard more reliable than a language.

And to survive, Eliot maintains, it must be an artistic language; "otherwise the spread of education will extinguish it."

Part of the unique artistic language of a culture is its musical language. Part of the function of fostering distinct cultures is to preserve unique regional modes of adapting to various nonhuman ecological systems. A multitude of cultures is a unique resource in a time in which, because of slim environmental resources, we are required to learn less standardized ways of habituating ourselves to extreme environments. To require extremely refined performance in one idiom as a measure of *educational progress* and to ignore quality performance in another idiom--say Sioux flute scales, which use non-Western scales--is not only to take away an expressive part of a culture. It is to leave a void that may have extreme effects on other aspects of a culture's adaptive or constructive capacities.

14. *Teacher Education in America* (Lincoln: Study Commission on Undergraduate Education and the Education of Teachers, in press), chap. 4. This publication, the final report of the Study Commission, Andrews 338, University of Nebraska, Lincoln, Nebraska 68508.

WHAT THE NAEP DELETED

It is possible to see how much meaningful educational progress can be measured by looking at what, as the National Assessment was constructed, was deemed offensive: for example, questions touching family, finances, references to specific minority groups, literary passages with sexual references, and questions dealing with birth control or religion. Exercises dealing with human rights were deemed offensive unless more exercises were added "dealing with... responsibilities in a free society." Deleted also were references to sex, unwed mothers, divorce, whisky, the FBI, the president, communism, and specific organizations such as the Ku Klux Klan and labor unions; references to violence or cruelty; exercises with *inappropriate* words or phrases, such as "sportive ladies leave their doors ajar;" exercises that might be interpreted as putting national

heroes or the police and other authorities in an unfavorable light; and exercises about the Civil War that suggest the North was better than the South. Senator Joe McCarthy, that demagogue and hate maker of the 1950s, is to be presented, according to the National Assessment, in a light that is neither too critical nor too favorable.[15]

Consider, if you will, a national assessment of educational progress that removes all of these areas of reality as "improper for questioning." What do you have left of the tough adult realities for which children must be educated, which they must learn to think about and deal with? In *The Changing Nature of Man*, Jan Van den Berg describes an inoffensive sex education lesson and asks, "What is left out of the picture of adult sexuality? Everything! The desire, the pain, the ecstasy--everything that makes adult sexuality meaningful." So it is here. The strictures against offensive material force the test to leave out everything--everything that makes education for competent adulthood worth struggling for in any specific culture or given any specific intellectual paradigm.

To measure national educational progress in terms that are tangential to survival in a specific community may not seem problematic so long as the assessment is only an assessment--something that somehow floats above the world of local decision-making and action in child rearing, above the pictures of "what is" and "how to live" that we give the young.

But National Assessment is not that in most areas. Local school districts readminister the Assessment and compile their own scores to judge how they are doing. System-wide curriculum decisions (and not at the local building level, as Keppel expected) are often based on the use of the assessment instrument. Keppel himself saw NAEP as a tool of national management. The Summer Institute report to the Office of Education by B.O. Smith and others,* which was connected to program development envisaged in connection with educational revenue sharing ("Renewal") and to followups to the Renewal effort, proposed using the National Assessment to determine priorities in schools using discretionary funds--for curriculum reform, teacher training, and management reform. Representative Albert Quie's (R-Iowa) proposal for Title 1, in many ways a humane proposal, would have allocated Title 1 funds on the basis of nation-wide criterion-referenced tests in math and reading, tests that almost surely would have come to be the National Assessment. As Guthrie and Emrick remark in their followup study on the Quie proposal, no *curriculum-neutral* tests can be created. Therefore, a national criterion-referenced test implies a national *course of study*.

> Similarly, in defining a domain, considerations of curricula are important. If the domain definition were (or attempted to be) inclusive of all current curricula, it probably would resemble current stand-

15. The list of exercises deemed offensive is from Finley, *National Assessment Approach*, pp. 42-46. Finley summarizes these items as "types of material considered offensive" by "lay reviewers."

*The Summer Institute on "Improvement and Reform of American Education" was an LTI-like effort mounted at the time "Renewal" was proposed. It continued after Renewal was rejected. The proposal to which I refer was included in a draft report of the institute, by Smith and others, which was circulated in the Office of Education.

ardized achievement tests. If the definition were curriculum-free, it might resemble an IQ test. If the definition were restricted to *dominant* curricula, it might be invalid for use with pupils whose educational experience is based on quite different curricula. This is to say, we do not yet fully understand the relative impacts of variations in curricula on educational attainment. It is most likely that these curricular variations produce corresponding differences in performance.

...It should be noted that a characteristic shared by most currently used domain-referenced and other criterion-referenced systems of measurement is that they were developed in conjunction with methods and materials of instruction (e.g., Hively et al., 1973). For this reason they are imbedded within particular curricula. Any nationwide application which attempts to take into account variations in curricula, therefore, appears fraught with problems, given the current state of the art.... The administrative logistics for a national assessment would be considerably different in cost and complexity than for 50 separate statewide or 16,000 districtwide assessments. And if multiple criteria or instruments are employed, the likelihood of methodological difficulties would increase by several orders of magnitude. The use of a single performance standard is strongly implied by (Quie's proposed bill). The presumed advantages of a single standard are that it would provide (1) a uniform definition of educational need, (2) an equitable basis for allocating program funds to states (supposedly overcoming the shortcomings of the multiple poverty criteria), (3) a good benchmark against which to evaluate program effect. Difficulties of this approach are in establishing the standard and in defending it in view of the enormous regional and local variation in curricula throughout the country....The creation of a National Commission on Educational Disadvantage, as suggested by Quie, for the purpose of determining the standard and developing a test would not alleviate any of the above developmental problems.[16]

Guthrie and Emrick also note that the Quie formula, as proposed in HR 5163, would represent a sharp break with American legal and congressional traditions.

HR 5163 continues to be under study in the House of Representatives and proposals similar to it have been put forward by Rep. Quie in recent months before the National Association of State Boards of Education; Rep. Quie is a strong proponent of parent control of the educational process. But that control rests not simply on having management tools through which parents can control the existing

16. John A. Emrick, James W. Guthrie, and others, *The Use of Performance Criteria to Allocate Compensatory Education Funds* (Menlo Park: Stanford Research Institute, 1974), vol. 1, p. 22.

process; it depends, too, on having management tools which reflect parent goals with respect to what the educational process ought to be. These goals will differ from culture to culture and from area to area. The National Assessment, as it is now funded and supported, makes no provision for the differentiation of goals from district to district or from culture group to culture group, and there is little prospect that this approach will be changed by congressional directive or by the technicians who make the Assessment, unless a serious effort is made to change it by parent and citizen participation groups.

It may be useful for such groups to ask of any National Assessment proposed in the future, "Were the Navajos asked? Did the Mennonites approve of it as representing their view of educational progress? What did the people of Harlem Prep, or any of the black or Chicano academies, say about it?"

In view of the overwhelming evidence that people learn to read and cipher, as well as to sing or draw or build, best in the idiom of their immediate environment and from the educational problems it poses for their own unique cognitive frames, and in view of the evidence that moving into the ecumenical world meaningfully requires that one first be at home in one's own world, it appears that the national interest would best be served by putting intense pressure on relating the evaluative process to:

- -local language and family or clan life;
- -local social structures and employment situations;
- -local natural environmental constraints;
- -the cognitive folk structures of the particular district and school in which the education is done.

If assessments are to be a tool at the building level, as the rhetoric implies, they have to permit principals, staff, and parents to relate the assessment to an area's unique culture, resources, problems, environment, and plans for the future. No assessment by itself can do that. That takes thought, and a quality too little appreciated in the language of present management--wisdom.

2

The Night Chant vs. *The Village Blacksmith:* A View of Literature

Perhaps the most graphic example of how explicitly the National Assessment is oriented toward Western culture is its literature instrument. The points I want to make are not examined here in depth, but they are telling.

The literature assessment has as its objectives that the child should:

(1) Read literature of excellence
  a. be acquainted with a wide variety of literary works
  b. understand the basic metaphors and themes through which man expresses values and tensions in Western culture
(2) Become engaged in finding meanings in evaluating works of literature
  a. respond to a work of literature
  b. find meaning in a work (literary comprehension)
  c. evaluate a work of literature
(3) Develop a continuing interest and participation in literature and literary experience
  a. be intellectually oriented to literature
  b. be affectively oriented to literature
  c. be independently active and encouraged by literature
  d. relate literary experience to life

First of all, goal one--*understand the basic metaphors and themes through which man expresses values and tensions in Western culture*--leaves out those one million Americans who come from Japanese and Chinese cultures; it leaves out nearly one million Native Americans, as well as that part of the literary heritage of Chicanos which depends upon Native American rather than Hispano sources; it also leaves out those African or Afro-Carribean works which express the independent civilizations and identities of black people.[1] The works which are mentioned for recognition are: *Little Red Riding Hood, Moby Dick, Little Bo Peep, The Turtle and the Hare, Sherlock Holmes, Alice in Wonderland, Winnie the Pooh, Charlotte's Web, Don Quixote, Casey at the Bat, The Village Blacksmith, The Charge of the Light Brigade,* and *Tom Sawyer.* The important literary characters are Noah, Samson, Adam, Venus, David, Gallahad, the Trojan Horse, Job, Daniel Boone, the Ugly Duckling, Rumplestiltskin, Paul Bunyan, Cupid, and Thor.

1. I can find in the instrument no reflection of the scholarship of Melville Herskovitz, Imamu Baraka, Roger Abrahams, Darwin Turner, and Jahnheinz Jahn, scholars who have traced the continuities between Africa and Afro-Carribean and Afro-American literary and artistic forms. Similarly, the scholarship on Native American, Chicano, and other than "mainstream" English literature appears to be ignored. More important, the test seems not to reflect an understanding of literature as something which, in many cultures, people construct to sustain themselves: to assist the crops, to order work routines, or to mediate some sense of the meaning of life. Rather literature is treated by the NAEP as something to "be acquainted with," "be oriented to," "appreciate," etc. Such relationships to literature are, at best, superficial: cf. Albert Lord, *The Singer of Tales* (Atheneum, 1960), and Northrup Frye, *The Educational Imagination* (Indiana University, 1965).

No one would knock the list of works that are proposed for recognition (save perhaps those bromide of bromide pieces, *The Village Blacksmith* and *The Charge of the Light Brigade*), but one wonders why only the sacred scriptures of Northern Europe and ancient Greece and Rome are brought in for recognition. Why not the Navajo Twins, or Quetzelcoetl and his epic, or Ananzi, or the sacred heroes whose thoughts are recorded in the *Bhagavad-Gita*? The answer is that Africa, South America, and native peoples in North America do not matter, according to the National Assessment. The ancient peoples who created *The Night Chant* lose out to *The Village Blacksmith*.[2]

Secondly, the literary analysis questions are often somewhat trivial. Children are asked, for example, to discover what the fourth line of the following poem should be:

Snow makes lightness where it falls
The bushes look like popcorn balls
And places where I always play

The alternatives proposed are:

Look like somewhere else today.
Look at me when I come to stay.
Look just the same as yesterday.

The correct answer is: "Look like somewhere else today." According to the National Assessment, only 40 percent of the nine-year-olds chose the correct answer; 42 percent said, "The places where I always play look just the same as yesterday." Is that to mean that there has been no progress at all? For most urban and many suburban children, the "places where I always play" are inside. They do "look just the same as yesterday" after a storm. For such children, an answer to a question phrased so abstractly has to be a guess.[3]

Thirdly, in several cases, the assessment is misleading. One test of literary skill is to identify puns in such sentences as "You have got a very good nose as noses run." The masters of the National Assessment say that if one says the pun is in the "run", one is wrong! "Noses run" is the true pun. Both words have double meanings! Another test of literary skill asks people to recognize the "similarities" between one passage from Shakespeare's *Caesar and Cleopatra*, which is a stage direction, and another passage, of dialogue, from Edward Albee's *The American Dream*. In yet another instance, the student is supposed to answer that William Carlos Williams' poem, "As the Cat," expresses the movements of a cat in the form of its tumble-tilt lines. As a matter of fact, people who know Williams' use of typography say that it is not at all clear that that kind of imitative form is what Williams is engaging in--form which makes punctuation and line-length move like a cat. Even if he is, it

2. I was the director of one of the 1960s Project English centers which threw out most of the old literary bromides (*Silas Marner*, *The Village Blacksmith*, etc.) established by the Committee of 10 Studies and related early twentieth century efforts making a standard curriculum. Project English centers had the effect of establishing a new canon: Aesop, *Charlotte's Web*, Greek myths, the Bible, *The Odyssey*, etc. The NAEP list seems to me to be a combination of the old and the new standard lists. Without wishing to derogate the accomplishments of the Project English centers, Francis Ianni and other federal officials connected to the establishment of the first centers indicated on numerous occasions that a "national curriculum" was not envisaged and that *many* centers were funded precisely to avoid "nationalization" of the curriculum. However, the NAEP's selection, for assessment purposes, of works which appear frequently in the various Project English curriculum guides, runs the risk of encouraging a new homogenization. The Project English centers also existed before establishment circles recognized local, folk, minority, or Third World literatures as potential school matter.

3. The limitations of literature assessment (or testing) which assumes some inevitability" in literature (as, say, in Haydn's music) may be inferred from June Downey's study, *Creative Imagination: Studies in the Psychology of Literature* (London: Kegan Paul, 1929), p. 183: "The reader was seen as tending to supply his own picture or background to an image." Nothing in later research (for example, Louise M. Rosenblatt, "The

is not clear that one should use National Assessments to determine whether students can recognize this particular subspecies of what O.K. Bouwsma has called the expression theory of art[4]--a theory of art based on the mistaken notion that art's power consists in its making some sort of formal imitation of what it talks about, its "expressing" it subject.

Much that is in the National Assessment's literature section makes a good deal of sense. The questions of what children think about the functions of literature and about the functions of what they have been reading, the questions which ask students to list titles of the books they have been reading, the categorizations of book titles that students list, the inclusion of opportunities for students to give reasons for teaching or not teaching literature--all of these can be helpful to teachers and students of culture in learning about young people's perceptions. Essentially, however, the academic part of the test regards literature as it regards music, as an academic exercise, not as a mode of expression which creates cultures and makes order in people's lives, not as the day-to-day, month-to-month, year-in year-out expression of a people trying to organize their lives. Work songs, folk songs, graffiti--according to the National Assessment--are not literature. Nor are the instructive and liturgical cycles of ancient peoples from all over the world.

When literature is literature for everybody--for no culture, "no peculiar way of thinking, feeling, and behaving"--no high claims can be made for it, and it is reduced to reading skills, technique, and memorization.[5]

Poem as Event," *College English*, *XXVI* (1964), 123-128; James Squire, *The Response of Adolescents While Reading Four Short Stories* (Urbana: NCTE, 1964), and Alan C. Purves, *Elements of Writing About a Literary Work* (Urbana: NCTE, 1968) contradicts the Downey finding. For the full literature, see Alan C. Purves and Richard Beach, *Literature and the Reader* (Urbana: NCTE, 1972).

4. Cf. O.K. Bouwsma, "The Expression Theory of Art," *Philosophical Essays* (Lincoln: University of Nebraska Press, 1965), pp. 21-50, esp. 47-48; cf. W.K. Wimsett, "Verbal Style", *The Verbal Icon* (Lexington: U. of Kentucky Press, 1967), pp. 201-217; cf. John Crowe Ransom, *The World's Body* (New York, 1938), pp. 95-97. One Williams' scholar whom I interviewed, Professor Mordecai Marcus, believes that Williams is engaging in imitative form in "As The Cat" (i.e., making the verse form move like a cat slinking) but regards the poem as an utterly trivial piece which should find no place in the NAEP.

5. For other reviews of the National Assessment of Literature, see Faye Louise Grindstaff and A. Muller, "The National Assessment of Literature: Two Reviews," *Research in Teaching English*, IX (Spring, 1975).

3

Rap Brown vs. Betsy Ross: A View of Writing

The same critique--that what is assessed is unsituated and meaningless--may be made of the National Assessment of composition. The objectives for the writing assessments are as follows:

(1) Write to communicate adequately in a social situation--letters, directions, formal notes, addressing envelopes, invitations.
(2) Write to communicate adequately in business or vocational situations, information and application forms, mail order letters, business invitations, formal letters.
(3) Write to communicate adequately in a scholastic situation: notes and announcements, and narrative, descriptive, and expository essays.
(4) Appreciate the value of writing (recognize the value of writing, write as a normal course of behavior, receive satisfaction from writing).

All but the fourth objective appear to be situated sorts of objectives. The instrument's objective appears to be to find out if a writer can do something in a specific setting. This accords nicely with the notion current in recent *ordinary language* philosophy that a word's meaning is its use; that, in asking what the meaning of a word is (or what the power of a word is in a particular situation), one really asks what that word or group of words (sentence) is doing as a tool in a particular *tool-shop*.[1] Yet, it is, as I see it, quite difficult, if not impossible, to simulate, through paper-and-pencil testing, the writing situations or sorts of situations which the writing assessment claims to measure.

For example, the *assessment* purports to measure a student's effectiveness in filling out the job application. Yet no job is in the offing when he fills out the assessment. Consider the difference, for a black student whose father and mother may have had a lifetime of casual, menial employment, between filling out a job application on a paper-and-pencil test which has no consequences and filling out a job application while sitting in a predominantly white employment office. As you write the application form, you remember that most of the people your age from your part of town are unemployed. Recession has led to the laying off of your Uncle Michael. In fact, recession-based

1. Cf. the exposition of the notion that "a word's meaning is its use," in O.K. Bouwsma's review of Ludwig Wittgenstein's *Philosophical Investigations*, reproduced in Bouwsma's *Philosophical Essays* (Lincoln University of Nebraska Press, 1965), pp. 197-199. Part of the difficulty of the Assessment is that it takes away from words the contexts in which we learn to use them.

early firings of blacks who do not have seniority have taken away a lot of jobs from people you know. Are you afraid? Do you perform better? I am not sufficiently a psychologist to know whether such pressures would make one fill out employment forms better or worse, but I do know that the experience of filling out a form where one's life may turn on what one does cannot be meaningfully reflected in job-form exercises in which one is "asked to write by examiners whom [one] does not know, told that [one's] teachers will not see [one's] writing, that it would not influence [one's] marks or academic futures, and [one] will receive no feedback at all on [one's] efforts." [2] In fact, one knows at the time of writing the assessment that no consequences will flow from what one does.

Writing which is inconsequential is unlike any other writing which one does. One commonly turns from the oral to the written language when important emotions, facts, records, or consequences are at issue. World War II student-pilots who were *assessed* where great issues were at stake were certainly assessed differently when they were asked midflight to handle the controls and save their own necks.

The difference between a situated and unsituated task may also be illustrated by the "Forest Fire" essay assigned to nine-year-olds. The essay assignment is as follows:

> Here is a picture of something sad that is going on in the forest. Look at the picture for a while. Do you see the forest fire? Write a *story* about what is happening in the picture. This is an important *story* because you want people to know about this sad *event*. [The picture depicts a forest fire with animals swimming across the river rapids to safety.] (Italics mine)

First of all, notice that the directions do not place the spectator and writer in a situation which approaches similitude. No one seeing a forest go up in smoke would go write a story--certainly not a child. "Help" is the cry. And no one seeing *a picture of a forest fire* would write "An important story because (he or she) wants people to know about this sad event." Children--at least the children I know--are likely to ask how the fire got started: "Did the animals get out? What do I do if I get caught like that? Will I die?" etc. Given the assessment's language (story? event?), it is not clear whether the child is being asked to pretend that the picture is a fictive picture which he is to frame with a story as in a children's book, or a *real picture* demanding that he respond to it--an important test of his capacity to use language in crises, to produce action.

2. John C. Mellon, *National Assessment and the Teaching of English* (Urbana: NCTE, 1975), p. 34. The notion that work which has no consequences will not be performed as if it were consequential is supported by contemporary psychology, as well as by contemporary philosophy.

HIGH-QUALITY AND MIDDLE-QUALITY WRITING

The student responses are just what one would expect given the unsituated nature of the task and the double focus of the directions. Consider two of the sample rated essays which have been released by the National Assessment of Educational Progress, one of which is rated as a typical high-quality essay and one which is rated typical middle quality* The *high-quality writing* is as follows:

> There was a fire in the woods one day. It was burning the trees down and burning the grass. It was a real big fire. There was a river near by but there wasn't anybody to put it out. There were rocks in the water. There were deer swimming across the river to get away from the fire. There was a raccoon on a rock trying to get across. The fire was getting bigger and bigger. The water was running down a small hill. (*85 words.*)

The student interprets the task as one that requires him to describe a sad event in the picture, eschewing fictive modes. The elements in the description are all also present in the picture, but there is little effort to see the picture as part of a continuum: no cause and effect, no impingement on the community, no beginning and no end.

---

*The procedure for holistic scoring is described as follows by Mellon (p. 23):

> Holistic scoring techniques have been extensively researched over the past twenty years, particularly by personnel of ETS in connection with essay exercises used in various College Board examinations. It is known, for example, that inter-rater reliability correlations (measure of the extent to which raters agree with one another on the rating assigned to a given essay) reach as high as .70 to .80 and above if raters are given special training sessions prior to their work.
>
> Equally high correlations are found between initial ratings and delayed reratings of a given essay by the same reader (after special training), thus verifying intra-rater consistency. Put more simply, we know that trained readers are consistent in their own overall-quality ratings and agree with the ratings of other readers about two thirds of the time. This is a far higher percentage than we initially thought, on the basis of earlier studies of judgments of writing ability, could ever be attained.
>
> Two key requirements of holistic scoring should be remembered if one is to fully understand the process. First, raters must judge individual essays

On the other hand, a so-called middle-quality essay reads as if its writer attached more importance to the direction to "write a story":

> Once upon a time there was a big forest there was 400 acres of trees and thousands of animals. But one day the forest caught a fire tree's was burning down and falling there was 2,000 firemen fight the fire, blazes was going 300 feet high in the air in two weeks the hold forest burned down and killed 500 hundred animals and 100 men and burned down 8,000 trees, and that was the end of the great forest fire. (*80 words*)

> relative only to the other essays in the group being rated, rather than to outside norms. Raters must use all rating categories a certain minimum percentage of the time, on the proven assumption that the general quality of any large collection of essays will distribute itself normally, that is, in familiar, bell-shaped fashion. Second, a rater is never permitted to base a rating entirely on any one aspect of the essay being read, no matter how outstandingly good or bad it may seem, but must always attend equally to all aspects--usually identified as content, organization, style, expression, and mechanics. This means, for example, that an otherwise mediocre essay may not be assigned the highest rating because the writer happened to succeed in deftly maintaining a humorous style. Nor may an essay, reasonably well-written in all other regards, be given the lowest rating just because the writer apparently lacks control over mechanical matters. Such is the logic of holistic scoring. Teachers may wish to reflect on the differences and similarities between this method of evaluating writing and procedures ordinarily used in the classroom.
>
> *Results of the Essay Exercises*
>
> Once the holistic scoring was completed, the assessors rank-ordered the essays, sorted by topic and age, and assigned the designations *high quality* to those at or above the 85th percentile of the overall-quality score distribution, *middle quality* to those below the 15th percentile. Although it is unclear what criteria of typicality they used, the assessors identified the following essays as *typical* of the three quality levels at each age:

Mellon then goes on to reproduce the essays assigned *typicality* at each level by the NAEP; they were the essays I used in this paper for most of my analysis. Notice that the criteria of quality are not listed and that there is no validation of a relationship between scorer-assessed quality and the objective features of *quality prose*, as these features might be described by linguists.

As one school of linguists and rhetoricians who grew out of the structuralist school would recognize, the middle-quality essay contains the *standard markers* for the beginnings and endings of *stories* intuitively understood by many elementary-school children: "Once upon a time," "and that was the end of the great forest fire."[3] Moreover, what is included ir between the markers is a story--not a simple effort to reflect the details of the picture at all. There were 400 acres of trees burned, thousands of animals frightened; 2,000 firemen came; there were blazes 300 feet in the air. Five hundred animals were killed and 100 men; 8,000 trees were burned down. The concentration of this story is on causes and effects: the forest catches on fire and this leads to the animal killing, the firemen fighting the blaze, their death, etc. The focus on causality, which goes with the sense of plot defined by Aristotle, and is present in almost all stories that pretend to be verisimilar, is present in this story.

Two pieces of writing--two *gestalts* or structures as to what is to be done--come into play because one child read the direction as a direction to write *a description of a picture* and the other *to write a story based on a picture*. The exact task at hand is ambiguous because the directions are ambiguous and because the task is unsituated. Structurally it makes no sense to compare the two pieces of writing. How is the judgment of *high quality* and *middle quality* on these two pieces, or on hundreds of others like them, to tell us whether our children, in the words of J.L. Austin, know how to do things with words.[4] The proper assessment would be that two different children, given ambiguous directions, did different but altogether plausible things with the directions.

Placing before students unsituated tasks leads to unsituated responses and some really silly data, to my mind. For example, a task given to 13-year-olds in the writing examination asks them to invite the mayor of one's city to speak at one's school. The invitation is meant to tell the mayor where the school is, suggest his topic, and set the date when he is to speak. The National Assessment indicates that such an answer as "I invite you to speak about your job to my class at Bryan Jr. High School whenever you can come", would be an acceptable answer. That may be an acceptable answer on a written test, but I know of no high school that would permit such an invitation to go out. Assemblies or classes are not called just any time. Moreover, few 13-year-old students are confronted with having to invite a mayor to speak at their schools, with suggesting his topic, and with discussing the choice of speaking dates with him. According to the assessment, only 27 percent of the students could do that task. Why? Because they were inadequate to the task or because it is a nontask?

As a matter of fact, most children in our society can probably concoct invitations that function for them.

3. Cf. Leslie Whipp, "Understanding Children's Writing," *Essays and Addresses on Composition* (Lincoln: NCDC, 1970), pp. 37-50, for a structuralist approach to children's stories which throws some light on the kind of story represented in the middle-level writing sample; cf. Thomas P. Klammer, "Some Recent Contributions to Tagmemic Analysis of Discourse" in Ruth M. Brend,, *Advances in Tagmemics* (New York: Elsevier, 1974), p. 380 (especially the discussion of aperture and denouement in relation to the middle-level sample quoted).

4. J.L. Austin, *How to do Things with Words* (London: Oxford Press, 1962). Austin's strictures on the meaninglessness of performative utterances, when uttered apart from the contexts where a social and linguistic convention declare, as it were, that they have meaning, is relevant to the NAEP composition test: cf. J.L. Austin, "Performative Utterances," *Philosophical Papers* (Oxford: Clarendon Press, 1970), pp. 233-52, and especially 237-239. The notion of performative utterances has been modified slightly in more recent philosophy, but Austin's statement is adequate for this context.

The roughest gangs know how to invite members to their rumbles, and the politest young ladies can get friends to their parties. Kids in the poorest, most illiterate classrooms can write notes to one another, in classrooms, which invite other kids to a fistfight or a kissing contest after school. I suspect that if a situation demanded it, many more than 27 percent of our children could write a formal letter of invitation to a mayor. After all, to have a task as a responsibility means, among other things, that if one doesn't know how to do it, one seeks assistance from other people in learning *how* to do it; one learns to routinize activities that one does over and over again; one seeks help from and imitates actions of other people who are doing similar kinds of things; one seeks help from friends or acquaintances. If one is uncomfortable with the writing of the invitational scheme, one does the invitation by talk, etc. Yet, the *Assessment* contains no validation of its notion that incapacity to write an acceptable invitation to the mayor in *Assessment* terms actually means that the same kids, in real life, couldn't get the job done.

One of the writing exercises, the "Famous Person" essay, appears to require no context. The direction in that exercise is as follows:

> Most of us look up to some famous person as a representative of the things we believe in and as the kind of person we would like to be. This person may come from any part of our society. For instance, we might admire Winston Churchill or Martin Luther King, Walter Schirra, or Mickey Mantle, Florence Nightingale, or Barbra Streisand. No matter where this person comes from or what kind of work he or she does, we can recognize such traits of greatness as determination, physical courage, ability to inspire others, and faithfulness to some worthy cause.
>
> Think about a famous person whom you admire. Select a particular admirable characteristic or quality of that person--such as Mickey Mantle's courage in the face of crippling physical handicaps or Florence Nightingale's determination to fight against strong governmental pressure. Write an essay about 200 to 250 words describing this characteristic or quality. Be sure to provide an illustration of it from the person's life. Try to show that the person is great at least partly because of this characteristic or quality.

What is called for here is essentially the rhetoric of admiration. One illustrative exemplary episode is meant to give particularity to the admiration which the essay is supposed to call into existence. No audience is described for the essay; it has no assigned truth value. It would appear to be a culturally neutral exercise. Yet, *truth* always has a context, and we are inclined to reward those

who praise the giants who belong to *our pantheon*. History, too, is a selective process.

What is troublesome here is that fine language and correct punctuation appear to come to be more important than *truth* or felt compassion in the *Assessment* judgments. Thus, when John Kennedy is praised in an essay rated *high quality* for his courage in the Cuban missile crisis (seen quite apart from the Bay of Pigs, American efforts to assassinate Fidel Castro, or any of the other events which we now know to have been part of the historical stream of which that crisis is a part), we may not blanch. But we do wonder what the language is supposed to be doing--what educational progress is being assessed? And progress toward what?

Where a 17-year-old, in a *high-quality* item, praises Sammy Davis Jr. as follows, we recognize the Fourth of July cliches marching in rows:

> Sammy knew that one day he would finally reach a place among his fellowman not as a black servant but as an equal brother to all of mankind. This goal he has already achieved. He has come, all the way, through dismal days, foodless days and shelterless nights. He knows what it is to suffer. He knows what it is to be hungry. But the fruits of labor can be so wonderful. Now he is a loved man among men.
>
> Sammy Davis Jr. has the ability to sing, dance, act and make friends. Sammy has dazzled every audience he has ever performed for with his God-given talent. With all of these qualities, how can anyone do anything but admire a man like that. (*183 words, Age 17*)5

5. Mellon, *Op. cit.*

On the other hand, the following prose about Martin Luther King is rated low quality:

> He was a famous man who did not believe in violence, just peace and brotherhood. He was a democratic person and a minister. He helped serve his country. He was a negro person who had a lovely family and when he died, or rather before he did, he stated, "I've reached the mountain top" ... He also made a longer speech, but I can't say it all, and his friend Jesse Jackson was also like him too and also, another great man; Reverend Abernathy. These three were great "non-violence" people. (*91 words, Age 13*)6

6. *Ibid.*

The prose here is unpolished. The version as originally punctuated is pretty unreadable. (So, for that matter, is a great deal of the academic prose I have read.) But somebody is home here. The values and perceptions are interestingly differentiated from the ordinary, run-of-the-mill observations in essays about King: democratic person *and*

a minister; not violence, just peace and brotherhood; he helped serve his country; "I've reached the mountain top!" The prose also comes from a person sufficiently open to say, "I can't remember it all"--"I admire King, Jackson, and Abernathy."

The papers are rated by ETS readers who are very ordinary middle-class people. They are, one suspects, like most of us, replicas of Auden's unknown citizen, who happen also to be English teachers. What are such people to do when they encounter two essays about Rap Brown and Betsy Ross side by side?

> The man I admire is an young man who goes by the name of Rap Brown. Rap is an young man who is trying to get the black people on their feet to try to go out and get what they should have. Many people do not agree with Mr. Brown on his ideas and they make fun and toy with him but he don't pay them any mind because he know they are nothing but a bunch of Uncle Toms. The only thing, I think that hurts Rap Brown is the idea he cannot get no kind of support from his fellow black brothers, the little support that he does get is not enough but still he fights on with what support he has and tries to get the black people in the United States right like the white man has. Many people do not agree with his ideas of violence's not one wants to be violent but when you talk and talk and not one listens, then there's only one thing left to do fight for what you think you should have. Rap Brown has been thrown in jail for what he has been doing but never the less he comes back stronger and stronger every time, till finally he is going to get the black people some of what they want.
>
> I admired Betsy Ross because she made the flag and when we were having a war and the men got sick she always helped them and tried to make them well. I also admired Helen Keller because she was so smart and could do so many things. Even though she was blind she learned how to read and write. When she grew up she wrote books for the blind and she even taught other blind people how to read and write.[7]

They give both the same rating, of course. Both are written off as middle-quality essays. Of such raters, one may ask, "What would they say should Catullus walk their way?"

THE FINICKY DETAILS OF WRITING

The second sort of analysis of writing in the English composition examination moves from the holistic to the finicky details of writing. Here, three kinds of procedures were

7. National Assessment of Education Progress, *A selection of Papers Collected During the 1969-1970 Assessment of Writing* (Washington: U.S. Printing Office, 1972), NAEP Report 10, p. 688.

used to characterize what was going on in the detail:

> *Procedure I:* computer counts of 24 quantifiable features of texts, such as average word and sentence lengths, number of each kind of punctuation mark, misspellings of certain words the computer had been programmed to look for, and so on;
> *Procedure II:* generalized prose characterizations of specific mechanical strengths and weaknesses of writing typical of the three quality levels at each age, prepared by panels of recognized experts;
> *Procedure III:* error counts of each of eight familiar types of errors, performed by teams of experienced English teachers. Two teachers working independently counted a given error type in each paper, and their totals were averaged.

Procedure I: The procedures raise some questions. In this kind of analysis, longer words are better than short words; Hemingway is poorer than Thorsten Veblen in such a scale, and the gibberish of someone like L. Ron Hubbard would top the scale. Longer sentences are also better than shorter sentences.[8] Seventeenth-century prose writers such as Sir Thomas Brown would rate tops, but Lancelot Andrews' short, repetitive sentences, if they appeared in the NAEP, would show a severe *lack of progress*. Similarly, Hemingway's style is apparently not syntactically progressive and Joyce's, in the most turgid sections of the conclusion of Bloom's day (*Ulysses*, pp. 627-698), would mark progress.

Procedure II: Specific mechanical *strengths and weaknesses* of writing typical of the three quality levels at each age were prepared by panels of recognized but unnamed experts. These *recognized experts* came to such conclusions as the following: Most 19-year-olds have limited competence in sentence construction and restricted vocabularies; the 13-year-olds who wrote the low-quality essays indicate "that the writers have no knowledge or understanding of the conventions of the written language;" "the poorest writing of the 17-year-olds was produced by teenagers who have no grasp of the conventions of the written language."[9] Yet the Martin Luther King paragraph quoted earlier, and treated as typical low-quality stuff, was clearly written by an author who has *some* understanding of the conventions of written language. (I quote the unedited paragraph):

> He was a famous man who did not believe in violence just peace and brotherhood. he was a democratic person and a minister. he help serve his country. he was a negro person who had a lovely family and when he died or rather before he did. he state I've reached the mountain top He also made a longer speech but i can't really say it all and his friend

8. See Sandra L. Stolsky, "Sentence-Combining as a Curricular Activity: Its Effect on Written Language Development," *Research in Teaching English*, IX (Spring, 1975), pp. 30-72, for some strictures on the use of sentence length as a measure of composition sophistication. Even Kellogg Hunt's "T-units" do not recognize the effective use of short syntactic units in some children's writing or in such writers as Hemingway or Lancelot Andrews.

9. The writers of the NAEP obviously did not understand the concept, "the convention of written language," in the way that most modern linguists do. If they did, they would have to admit that the *low-quality* essays showed considerable mastery of, say, the graphemic structure of written English. For a short statement as to linguistic perspectives, see David Crystal and Derek Davy, *Investigating English Style* (Bloomington: Indiana University Press, 1969), pp. 16-20; 68-71.

Jessy Jackson was also like him too and also another great man. Reverend Avernathy there were three great non-violence people. (*91 words*)

Periods are generally marked; words are spelled conventionally for the most part. Though the student does not write as I do, he/she obviously commands some of the devices. Capitalization is not conventional, but there is at least some beginning movement toward the use of conventional capitalization: "Jessy Jackson," "Reverend Avernathy," "He," etc.

It should be observed that editing problems such as appear in the so-called low-quality prose appear also in the writing of college professors. For the last decade, I have edited the Regent's Critics series for the University of Nebraska Press. The following is a list of editorial problems fairly typical of those which appeared in the writing of subordinate editors presenting a volume to be published; on the left side, I give the text as written and on the right the revision which the editor suggested:

[*Sense*]: obvious form of the/obviously derivative from
romantic myth the romantic myth...

[*Spelling*]: metomorphose/a metamorphosis as Vigny gave
them verbal representation

[*Pronoun*]: he gives us/Sainte-Beuve gives us

[*Verb Choice*]: is an aesthetic/has an aesthetic

[*Punctuation*]: ,the analogue of/;Bossuet's mind is the
analogue of Bossuet's
eloquence

[*Pronoun*]: he sees/Sainte-Beuve sees

[*Connective*]: of more recent/of such more recent
critics like critics as Maurice...

[*Article*]: Such affinities/The affinities

[*Connective*]: today may do/today, affinities which I
have described, may do

[*Sentence Structure*]: They would do/These affinities would be
even even more attractive if
Sainte-Beuve had made it
a practice

[*Pronoun*]: As he grew/As Sainte-Beuve grew older, his
older earlier preference...

[*Neuter Pronoun*]: school of L.H. .../school of L.H. ... which whom

These are the problems in two pages of prose. Misspellings, incorrect punctuation, confusing syntax, and pronoun reference problems appear frequently in my notes on the text/--and I quote from my notes on one of the cleanest, best written books which I have ever received.

Almost all of the prose that I published as editor of the Regents Critics went through several hands before seeing print. In fact, very few professional writers publish their work without having friends and editors review the writing, and often do so several times before publication. More to the point, few people write for unsituated demand in the way that the National Assessment asks that they write. Anyone who wishes to see what the spontaneous writing of people in this country was like 40 years ago should look at C.C. Fries' American English Grammar;[10] to find out what it is like today, one need only clear the wastebaskets of first drafts of letters in a businessman's office; then let the judgment of educational progress be made. The fact is that editing is part of the prose process, and a most important part. This opportunity for reflective editing is denied by the National Assessment.

Procedure III: In view of our lack of information about what appears in professional writing,[11] the quantification of *errors*--incorrect punctuation and capitalization, fragments, run-ons, awkward constructions, agreement problems and incorrect word choices [Procedure III]--is irrelevant to a progress assessment. For one thing, the NAEP writing was not done for publication. It contains the *problems* that ordinary English writing commonly contains. Add to that--the fact that the NAEP measure is based on no accurate descriptions of the practice of experienced or professional writers or the editing assistance offered them--the fact that neither does it draw on descriptions by ordinary citizens of how they actually use written language, and you have an assessment deprived of the tools of judgment. The *error counts* will only assist empty-headed, *back-to-basics* people who have no real information about how the English language is used in context; they will never assist a teacher who, or a school that, wants to help a student write.[12] Even as the NAEP sample never asks whether anybody is at home or where home is, it is likewise not based on any research into what someone who had "a local habitation and a name" would do when asked to write.

10. C.C. Fries, *American English Grammar* (New York: Appleton-Century, 1940).

11. During the Project English period, some efforts were made to characterize the actual practice of professional writers of English; this work was carried ahead by Francis Christensen, Richard Young, Alton Becker, Kenneth Pike, and others. For notes on how this sort of description might be done, see Francis Christensen, *Notes Toward a New Rhetoric* (Harper & Row, 1967), and Richard Young, Kenneth Pike and Alton Becker, *Rhetoric: Discovery and Change* (Harcourt-Brace, 1970). The effect of the Assessment's pronouncements on writing (most of which do not grow out of a rigorous analysis of the test and its responses conducted by linguistic rhetoricians) is likely to force schools to adopt increased work in traditional grammar and routine writing practice.

12. These observations from recent surveys of in-depth research are in order: "statistical and nonexperimental studies [by all recent researchers] failed to show a significant relationship between grammatical knowledge and writing ability;" "the research overwhelmingly supports the contention that instruction in formal grammar is an ineffective and inefficient way to help students achieve proficiency in writing;" *linguistic grammars* have shown limited results save in sentence-combining exercises" (transformations, however, are not looked at in the NAEP); "writing alone does not teach writing." Cf. A. Stephen Sherwin, *Four Problems in Teaching English: A Critique of Research* (Scranton: International Textbook Co., 1969), p. 168.

4

The National Assessment of Reading: Horsepower Without Horse Sense is Fatal

The reading section of the National Assessment for Educational Progress includes the following objectives, in terms of which assessment is to be made of 9-, 13-, and 17-year-olds:

1. Understanding words and word relationships (literal comprehension of isolated words, phrases, and sentences);
2. Graphic materials (the comprehension of the linguistic components--drawings, signs, labels, charts, maps, graphs, and forms);
3. Written directions (comprehension of directions, plus ability to carry them out operationally);
4. Reference materials (comprehension and knowledge of indices, dictionaries, alphabetizing, and TV listening formats);
5. Gleaning significant facts from passages (comprehension and, to a limited extent, recall of literal content and the context of a larger reading passage);
6. Main ideas and organization (ability to abstract upwards from the sentence-by-sentence content of a passage and recognize the main ideas and organization features);
7. Drawing inferences (ability to reach conclusion not explicitly stated in the passage, in most instances relying only on information given, and in a few cases on knowledge unrelated to the passage);
8. Critical reading (ability to recognize author's purpose, and to understand figurative language and literary devices).

Notice that none of the objectives has to do with reading in terms of the specialized vocabulary of a profession or vocation, possibly because the Assessment, being national, cannot easily look at reading for all vocations and cannot look at the different ranges of vocations that would be present in any specific community or culture. The materials chosen for specific understanding under the objectives for graphic and reference materials are the kinds which go along with a highly mobile, urban population: signs, charts, maps, graphs, TV listing formats, indices. They preclude such specialized items as comprehending directions for assembling a specific kind of mo-

Stolsky, in the study cited previously, observed that sentence-combining activities which are directed toward teaching students a described professional or situated syntactic reportoire may be of some use. She also observed that "the special and typical characteristics of mature written language structures have never been systematically explored from a pedagogical perspective (p. 66)." It is difficult to see how an Assessment can be made prior to the creation of such a research base.

tor, comprehending directions for determining the sex of baby chicks, reading a book giving directions as to how to raise a garden.[1] Nor do they include reading exercises in any non-English language, although several million of our people, including some children in the schools, are more comfortable in reading a language other than English than they are in reading English.[2]

Since the questions are not environment-specific, they have to attempt to assess *generally* needed skills, which, in turn, require a rhetorical format. It appears that the questions with which the students had the most difficulty were, in fact, those that involved the use of a rhetorical format not in universal use in society. Among the 17-year-olds, for instance, the recipe format bothered 38 percent; the graph format bothered 26 percent (the graph was presented with little of the textual material that conventionally accompanies a graph), and a report card format bothered 20 percent (apparently it was assumed that students would make an elaborate referencing between several columns, back and forth, in a complex report card form). The fact that report cards, in the main, are made for parents, and that, in many school systems, they are more discussed with parents than students, may enter into the difficulty that the 20 percent had in reading them.

Many people had difficulty reading some items in a TV guide (about 20 percent of both the adult control group and the 17-year-old student group). But, then, almost 5 percent of American households do not even include TV sets.[3] And, of these, it might be reasonable to suspect that less than 50 percent get a TV guide. Nor is the reading of a TV guide a skill taught in most schools. If one neither has a television set nor bothers to get a TV guide there is very little reason to know how to read one. Yet, the Assessment of Educational Progress was made in terms of whether one can read the technical language, peculiar to television guides, in such a table as the following:

EXHIBIT 7. Exercise R413 Stimulus

Here is part of a TV guide you might find in a newspaper. After reading it, answer the questions which follow it.*

2:00 p.m.	[2]	Super Mutt - Cartoons
	[4]	Baseball [Color] Teams to be announced. (Runs to 4:00, followed by Baseball Scoreboard.)
	[6]	Top Cat - Cartoons

*Adapted from *TV Guide*, with permission.

1. The NAEP Reading Test appears to be suited to meet the goal urged by the NIE Study Group on Linguistic Communication: "We must measure the reading levels that are required to handle specific kinds of reading materials that are used on the job, in the home, and in other life situations, and measure progress against these stable markers rather than the rubber-rulers of norm-referenced standards." However, the assessment is so general that no specific vocation's reading repertoire is included. Moreover, as the NIE Task Force observed, the evidence as to reading levels needed in the various vocations is "scattered." Cf. George A. Miller, *Linguistic Communication: Perspectives for Research* (Newark, N.J.: IRA, 1974), pp. 13 and 15.

2. Title VII ESEA officials in 1975 placed the number of limited-English and no-English speaking children in the U.S. at between 8 and 12 million. Cf. *Teacher Education in The United States: The Responsibility Gap* (Lincoln, Neb.: University of Nebraska Press, 1976), p. 129. *The Condition of Education: 1978* (U.S. Gov't. Printing Office) places the figure lower, but still in the millions.

3. This information was furnished me from standard TV encyclopedias by John Flower of Channel 12, Nebraska ETV.

[7] Movie - Mystery
"Master Sleuth" (1945) Master detective (Rob Johnston) and sidekick (Pat Morgan) are on the trail of a deadly escaped convict. Sue Jones, Mort Roberts (90 min.)

2:30 p.m. [2] Children's Variety (Repeat)
Today the show goes to Detroit to watch cars being assembled. (60 min.)

[6] Visit the Zoo (special)
Famous San Diego Zoo is toured.

3:00 p.m. [6] Music Beat.
Jay Nickels hosts an hour of popular music of local groups.

The student is supposed to read the item listed for 2:30 p.m., for instance, and know that the word (Repeat) in that item means that "the program is being run for a second time." Since I do not own a television set and don't read TV guides, I would not have answered this item correctly had I taken the assessment.

Very obvious materials can give students a great deal of trouble. The following question was asked as part of the reading examination:

> "Horsepower without horse sense is fatal." Where would you probably see this sign? On a gymnasium floor? At a racetrack for horses? On a highway? In a grocery store? I don't know?

The correct answer is "On a highway." It was given by 23 percent of the 9-year-olds, 45 percent of the 13-year-olds, and 76 percent of the 17-year-olds. The Assessment reasoned that the people who did not answer correctly did not know the meaning of horsepower and, therefore, did not associate the sign with the highway warning. It is also possible, however, that some of the so-called inference answers were given by students who, seeing the pun on horsepower and horse sense, assumed very different contexts from those assumed by the examiners. That 76 percent of the 17-year-olds answered "correctly" suggests that most students learn to read the sign as a highway warning by the time they learn to drive.[4] There is no context that I can think of in which my 12-year-old daughter needs to know how to act on the warning: "Horsepower without horse sense is fatal."

Problems which occur when students are encouraged, or required, to supply a meaningful context for a question may be illustrated by the following passage about a South American village:

4. The fact that "reading levels" requisite for most occupations or for normal consumer activities remain undertermined undoubtedly handicaps the NAEP reading assessment's effort to assess whether students are achieving "the general reading competence necessary to the citizen." Cf. the report of the Study Group on Linguistic Communication, NIE, *Linguistic Communication: Perspectives for Research* (IRA, 1974), pp. 12-14.

Village of Nayon

Until about thirty years ago, the village of Nayon seems to have been a self-sufficient agricultural community with a mixture of native and sixteenth century Spanish customs. Lands were abandoned when too badly eroded. The balance between population and resources allowed a minimum subsistence. A few traders exchanged goods between Quito and the villages in the tropical *barrancas*, all within a radius of ten miles. Houses had dirt floors, thatched roofs, and pole walls that were sometimes plastered with mud. Guinea pigs ran freely about each house and were the main meat source. Most of the population spoke no Spanish. Men wore long hair and concerned themselves chiefly with farming.

The completion of the Guayaquil-Quito railway in 1908 brought the first real contacts with industrial civilization to the high inter-Andean valley. From this event gradually flowed not only technological changes, but new ideas and social institutions. Feudal social relationships no longer seemed right and immutable; medicine and public health improved; elementary education became more common; urban Quito began to expand; and finally--and perhaps least important so far--modern industries began to appear, although even now on a most modest scale.

In 1948-49, the date of our visit, only two men wore their hair long; and only two old-style houses remained. If guinea pigs were kept, they were penned; their flesh was now a luxury food, and beef the most common meat. Houses were of abode or fired brick, usually with tile roofs, and often contained five or six rooms, some of which had plank or brick floors. Most of the population spoke Spanish. There was no resident priest, but an appointed government official and a policeman represented authority. A six-teacher school provided education. Clothing was becoming citified; for men it often included overalls for work and a tailored suit, white shirt, necktie, and felt hat for trips to Quito. Attendance at church was low and many festivals had been abandoned. Volleyball or soccer was played weekly in the plaza by young men who sometimes wore shorts, blazers and berets. There were few shops, for most purchases were made in Quito, and from there came most of the food, so that there was a far more varied diet than twenty-five years ago. There were piped water and sporadic health services; in addition, most families patronized Quito doctors in emergencies.

The crops and their uses had undergone change. Maize, or Indian corn, was still the primary crop, but very little was harvested as grain. Almost all was sold in Quito as green corn to eat boiled on the

cob, and a considerable amount of the corn eaten as grain in Nayon was imported. Beans, which do poorly here, were grown on a small scale for household consumption. Though some squash was eaten, most was exported. Sweet potatoes, tomatoes, cabbage, onions, peppers and, at lower elevations, sweet yucca and arrowroot were grown extensively for export; indeed, so export-minded was the community that it was almost impossible to buy locally grown produce in the village. People couldn't be bothered with retail sales. (From "The Village in an Industrial World," *Scientific Monthly*, August, 1953.)

A question asked concerning this passage, and the percentage of answers attached to each possible response, were as follows: By 1948, the village of Nayon was (a) a self-sufficient village (28 percent, 13-year-olds; 15 percent, 17-year-olds); (b) out of touch with the outside world (6 percent, 3 percent); (c) a small, dependent portion of a larger economic unit (14 percent; 23 percent); (d) a rapidly growing and sound social and cultural unit (48 percent, 54 percent); (e) don't know/no response (4 percent and 5 percent).

Answer (c), "a small, dependent portion of a larger economic unit," was supposed to be the correct answer. The two items that most confused the test-takers were whether Nayon, as described in the passage, is a small dependent portion of a larger economic unit or a rapidly growing and sound social and cultural unit. The passage sufficiently eliminates some of the detractors. When the village has to import food such as corn and go outside for medical service, it is hard to describe it in any meaningful terms as self-sufficient, and only 15 percent of the 17-year-olds saw it so. The notion that the village is now out of touch with the outside world is pretty clearly what the author is trying to show that it is not, and only 6 percent of the 13-year-olds and 3 percent of the 17-year-olds were confused by that distractor. But the fact that the date given to assessment students in the early 1970s *in the prose passage* refers to the persistence of the older order "30 years ago" [i.e. 30 years before the writing in 1953, around 1923, and not 30 years before 1970 (1940)] is confusing, particularly in view of the other use of the 1948 date to mark the new era. Students who had little time to coordinate dates and footnotes may have seen 1948 and "thirty years ago" (mistakenly projected from 1970) as about the same time. Thus, they may have been confused about which paragraphs describe the new and which the old order in Nayon. If they were, the third and fourth paragraphs of the passage, describing change (the new railroad to Guayaquil-Quito, the abandonment of feudal social tradition, improvements in public health and medicine), may easily have suggested to students that the Nayon of 1948-49 was a "rapidly growing and sound social and cultural

unit," particularly in view of the fact that the key phrases in the distractor have no definition. Neither "rapidly growing" nor "sound" nor "social and cultural unit" are defined. The student who understands that the focus of the question is on a particularly academic anthropological and economical definition of "unit" will answer correctly, if he understands the passage, but I do not believe that a knowledge of such technical language is a test of capacity to read. Nor is the rapid correlation of dates in a complex passage.

Not only is English used on speakers of all languages, putting some at a disadvantage, but where speakers of English are tested, there is no control over registers, dialects, or styles which would permit some sort of analysis as to what the schools are doing wrong. Black students are said not to read well. Which black students? Those who have been educated in homes where mainstream dialects are spoken? Those brought up in homes where such dialects are not spoken? Gullah-speaking blacks? Blacks having parents who never read? How do poverty, nutrition, even an oral tradition, enter in? The Assessment solicits no such information at that level of social or linguistic nicety.[5]

One final caveat: 35 percent of the nine-year-olds were unable to read and confirm the *true* contents of this

---

[5]See Kenneth R. Johnson, "Black Dialect Shift in Oral Reading: How Black Dialect-Speaking Children Impose Their Grammatical Features on Standard English Texts During Oral Reading," report of USOE research, read at the 1976 National TESOL Conference, for research relating dialect information to evaluation of "correct reading." Part of the abstract of the research is as follows:

> During oral reading, Black dialect-speaking children often make the following typed or apparent reading miscues: (1) substitutions; (2) omissions; and (3) insertions. However, if their miscues are consistent with the grammatical features of Black dialect--in other words, their apparent miscues are direct translations from standard English to Black dialect with no loss of comprehension--these types of miscues cannot be treated as "normal" or "regular" oral miscues. Instead, this phenomenon is here labeled "dialect shift."

The classification of reading data by "race of student" as opposed to "language and culture" of student conforms to no canons of linguistic investigation with which I am familiar. The failure to collect dialect and second language data militates against the NAEP making a contribution to answering the research and policy questions with respect to dialect and reading posed by the NIE Study Group.

sentence: "Bubble gum that never loses its sugary flavor would stay sweet for a long time." A sympathetic and quiet critic of the Assessment observes that the failure of these nine-year-olds to do so stems from one or more of the following factors: inability to decode the letters that comprise "bubble" (or "sugary," or any other of the pronounceable word sounds); lack of knowledge of the word meaning, the concept associated with each word sound; inability to compile the words once read into a recognizable syntactic structure; or failure to recognize the syntactic functions of the negative of time, "never," or of the conditional modal auxiliary "would." The critic goes on: "Some reading problems result from not knowing the meaning of written words one can pronounce, others from not being able to associate the written forms of the words with meanings and sounds one otherwise knows, and still others from inability to recognize the syntactic features of a sentence in which the words occur. The important point is this, that until these three factors are isolated and controlled in the items of a reading test or assessment, it is wrong to believe that the test or assessment is measuring awareness of higher level features of discourse or the ability to perform certain cognitive operations. But the NAEP reading exercise neither provides separate information on word-decoding, word knowledge, and syntactic processing, nor does it control these variables in seeking to measure discourse-reasoning and cognitive thinking. As a result, from the point of view of someone interested in special studies of reading comprehension, the insights to be derived from the reading assessment are rather limited."[6]

Indeed, they are.

6. John C. Mellon, *National Assessment and the Teaching of English* (Urbana: NCTE, 1975), p. 69.

5

Toward an Alternative Assessment

Thus far, my argument has pointed out that:

1. The National Assessment was created through a politics of artificial consensus which runs counter to our legal tradition with respect to local control of the schools, and flies in the face of many of the formal and informal educational practices of the multitudes of cultures and groups which make up the country;

2. As a consequence, the assessment schedule reflects artificial consensus positions, positions which include neither the concerns of unusual cultures nor those of the scholarly societies made up of members of those cultures, nor the unusual scholarly career concerns of what Thomas Kuhn would call "the constructors of new paradigms" (e.g. Chomsky).

3. Because the assessment assesses things which are so general, it either eschews policy consequences--in which case it is a waste of money--or it has policy consequences which are likely to push cities, states, or regions away from reliance on their own norms, assessments, and judgments as to how the education of children is to be conducted.

4. The assessment tool, in its effort to reflect "general" notions as to what the expressive arts and language (reading, composition, literature) should be and do, overlooks or blurs the specific and quite various meaningful functions that forms of each performs in its own vocational, cultural, or community context.

5. The result is an assessment which is fairly useless for policy purposes and which, in language areas, neither bases itself firmly in past research nor points the way to future research of a meaningful sort.

This is not to say that some individual findings (e.g. the reading habits of high school youngsters) may not be exceedingly interesting. It is not to say that the women and men who wrote and reported the assessment were not intelligent, honest, hardworking, and well-meaning. It is to say that the assessment had the wrong base and asked the wrong questions, if we are to learn to assess so as to educate children in Crystal City, Texas; Chinle, Arizona; Chicago, and Kenawha County, West Virginia.

A different sort of assessment would ask in each neighborhood, and in each culture and subculture, that we (parents, supervisors, teachers, other adults) know as much as we can about our children and about what it takes to be an adult in our culture, given the land or natural (physi-

cal) environment in which one has grown to maturity. Then we might together engage in a constant effort to determine "Is the kid growing up right? Does he or she know what needs to be learned, if we are to be what we want to be?" Using whatever good judgment we can, we could act on such conversations. This is not a scientific way to assess. The assessment would come out as "just talk" about what we, as a community, should do about the next generation. But, at least, the talk and the assessment would not produce a judgment that had the appearance of science or objective information and none of its essentials. The kind of model I am talking about could be developed in terms of day-to-day, week-to-week, month-to-month assessments of the growth of children, individually and in groups, toward goals articulated or less directly envisaged or expressed by the fostering community.

Such assessment may be a hard job. I take comfort in two observations of Courtney Cazden's with respect to when children's language (talk and other language skills) cannot be assessed and when it can. I quote at length:

> [First] Shapiro (1973) has documented the failures she encountered in trying to assess behaviors of young children previously observed in a primary-grade classroom following a Bank Street model of education--in an artificial context where those same behaviors had no functional value for the child and were now requested only in response to a test-like question of some unfamiliar adult.
>
> [Cazden then advocates assessment in the context of a direct, concentrated, personal encounter between one child and another, and between teacher and student in the school.] In contrast to the more contrived situations we call *tests*, the most important characteristic of concentrated encounters is that they are condensed forms of familiar interaction experience. They represent our best examples of teaching encounters and are as close as possible to them in setting, participants, and topic. *But for assessment purposes, they are more focused by teacher direction and involve a smaller than usual group of children so that the participation of each child is maximized.*
>
> Familiarity is the key precisely because the situational influences on speech are so powerful that it is difficult, if not impossible, to get a young child to transfer language skills he has demonstrated in a natural situation to some more contrived situation in which we wish to elicit them on demand.
>
> The important issue is not whether the teacher or some separate *experimenter* conducts the encounter. I think

the teacher is the most advisable adult both for getting the best from the child (unless we wish particularly to assess his ability to relate to strangers), and because the teacher will herself obtain first-hand information on the child's behavior which should be useful in planning future language experiences.

The more important issue is the degree of freedom or standardization of the teacher's behavior. If the teacher's behavior has been *scripted* as it is in the traditional tests, then no information on her *typical* behavior is available; the information on child behavior is more comparable across teachers, but full interpretation of that behavior would depend on some notion of the representativeness of the task in the ongoing life of the school. Does the script fit the norms of interaction in the classroom setting? How often are opportunities for practice of the requested behavior available and taken up by each individual child? These are both questions of process-product relationships.

If the teacher's behavior is not scripted, then the encounter can be well used to describe typical teacher behavior, but information on individual children will be less clear. The teacher may bring out the best in a child or not. The child's responses would at least have to be weighted in terms of the opportunities provided in the encounter for particular kinds of response. In the *Schools Council Pre-School Language Project* (1972), Joan Tough describes both the value of nonscripted encounters and their limitations:

As conversation progresses, decisions will have to be made about the direction in which the child's interest seems to lie, and what aspects of the experience are likely to have the kind of meaning which he will be ready and perhaps eager to express... We must remember, however, that in our examples the teacher asked different questions and we cannot know whether it was the questions which restricted the answers... or whether this would have been the character of the children's responses if the questions had been more open.

*Perhaps if the same child participates in several such encounters, then the context-dependent evidence from each can be accumulated across encounters to yield corroborative evidence for a more complete picture of each child.**

*Courtney Cazden "Concentrated vs. Contrived Encounters: Suggestions for Language Assessment, *The Urban Review*, Vol. 8, No. 1 (1975), pp. 31-33.

It is this kind of tentative, context-specific assessment, particularly if it is designed by, or with the help of, parents and/or the extended family, which is, I think, likely to assist the sense of responsibility of both the local area in determining whether it is doing what it ought in the education of children and, by accretion, the nation in developing solid, community-based work.

Also available as part of the North Dakota Study Group on Evaluation series:

Observation and Description: An Alternative Methodology for the Investigation of Human Phenomena
Patricia F. Carini

Alternative Evaluation Research Paradigm
Michael Quinn Patton

An Open Education Perespective on Evaluation
George E. Hein

A Handbook on Documentation
Brenda Engel

The Teacher Curriculum Work Center: A Descriptive Study
Sharon Feiman

The Word and the Thing: Ways of Seeing the Teacher
Ann Cook and Herb Mack

Testing and the Testing Industry: A Third View
John Williams

Special Education: The Meeting of Differences
Steven D. Harlow

Psychological Effects of Open Classroom Teaching on Primary School Children: A Review of the Research
Robert A. Horwitz

Children's Interactions in Traditional and Nontraditional Classrooms
Sylvia Ross, Herbert Zimiles, and David Gerstein

Developing Hypotheses About Classrooms from Teachers Practical Constructs
John Elliott

Single copies $2, from Vito Perrone, CTL
U. of North Dakota, Grand Forks, N.D. 58202